Chapter 8

Distributed Sorting

"Indeed, I believe that virtually *every* important aspect of programming arises somewhere in the context of sorting [and searching]!"

– Donald E. Knuth, The Art of Computer Programming

In this chapter we study a classic problem in computer science—sorting—from a distributed computing perspective. In contrast to an orthodox single-processor sorting algorithm, no node has access to all data, instead the to-be-sorted values are *distributed*. Distributed sorting then boils down to:

Definition 8.1 (Sorting)**.** *We choose a graph with n nodes $v_1, \ldots, v_n$. Initially each node stores a value. After applying a sorting algorithm, node v_k stores the kth smallest value.*

Remarks:

- What if we route all values to the same central node v, let v sort the values locally, and then route them to the correct destinations?! According to the message passing model studied in the first few chapters this is perfectly legal. With a star topology sorting finishes in $\mathcal{O}(1)$ time!

Definition 8.2 (Node Contention)**.** *In each step of a synchronous algorithm, each node can only send and receive $\mathcal{O}(1)$ messages containing $\mathcal{O}(1)$ values, no matter how many neighbors the node has.*

Remarks:

- Using Definition 8.2 sorting on a star graph takes linear time.

8.1 Array & Mesh

To get a better intuitive understanding of distributed sorting, we start with two simple topologies, the array and the mesh. Let us begin with the array:

Algorithm 8.3 Odd/Even Sort

1: Given an array of n nodes $(v_1, \ldots, v_n)$, each storing a value (not sorted).
2: **repeat**
3: Compare and exchange the values at nodes i and $i+1$, i odd
4: Compare and exchange the values at nodes i and $i+1$, i even
5: **until** done

Remarks:

- The compare and exchange primitive in Algorithm 8.3 is defined as follows: Let the value stored at node i be v_i. After the compare and exchange node i stores value $\min(v_i, v_{i+1})$ and node $i+1$ stores value $\max(v_i, v_{i+1})$.
- How fast is the algorithm, and how can we prove correctness/efficiency?
- The most interesting proof uses the so-called 0-1 Sorting Lemma. It allows us to restrict our attention to an input of 0's and 1's only, and works for any "oblivious comparison-exchange" algorithm. (Oblivious means: Whether you exchange two values must only depend on the relative order of the two values, and not on anything else.)

Lemma 8.4 (0-1 Sorting Lemma)**.** *If an oblivious comparison-exchange algorithm sorts all inputs of 0's and 1's, then it sorts arbitrary inputs.*

Proof. We prove the opposite direction (does not sort arbitrary inputs $\Rightarrow$ does not sort 0's and 1's). Assume that there is an input $x = x_1, \ldots, x_n$ that is not sorted correctly by the sorting algorithm. Then there is a smallest value k such that the value at node v_k after running the algorithm is strictly larger than the kth smallest value $x(k)$. Define an input $x_i^* = 0 \Leftrightarrow x_i \leq x(k)$, $x_i^* = 1$ else. Whenever the algorithm compares a pair of 1's or 0's, it is not important whether it exchanges the values or not, so we may simply assume that it does the same as on the input x. On the other hand, whenever the algorithm exchanges some values $x_i^* = 0$ and $x_j^* = 1$, this means that $x_i \leq x(k) < x_j$. Therefore, in this case the respective compare-exchange operation will do the same on both inputs. We conclude that the algorithm will order x^* the same way as x, i.e., the output with only 0's and 1's will also not be correct. □

Theorem 8.5. *Algorithm 8.3 sorts correctly in n steps.*

Proof. Thanks to Lemma 8.4 we only need to consider an array with 0's and 1's. In this array, let j_1 be the node containing the "rightmost" 1, i.e., j_1 is the highest-index node from $\{v_1, \ldots, v_n\}$ with value 1. If the index r of node $v_r = j_1$ is odd (even), the value 1 of this node will "move to the right" in the first (second) step. In any case it will move right in every following step until it reaches the rightmost node v_n. Let j_k be the node with the kth rightmost 1. We show by induction that j_k is not "blocked" anymore (constantly moves until it reaches destination!) after step k. We have already anchored the induction at $k = 1$. Since the 1 at node j_{k-1} moves after step $k-1$, j_k gets a right 0-neighbor for each step after step k. (For matters of presentation we omitted a couple of simple details.) □

Remarks:

- Linear time is not very exciting, maybe we can do better by using a different topology? Let's try a mesh (a.k.a. grid) topology first.

Algorithm 8.6 Shearsort

1: We are given a mesh with m rows and m columns, m even, $n = m^2$.
2: The sorting algorithm operates in phases, and uses the odd/even sort algorithm on rows or columns.
3: **repeat**
4: In the odd phases $1, 3, \ldots$ we sort all the rows, in the even phases $2, 4, \ldots$ we sort all the columns, such that:
5: Columns are sorted such that the small values move up.
6: Odd rows $(1, 3, \ldots, m-1)$ are sorted such that small values move left.
7: Even rows $(2, 4, \ldots, m)$ are sorted such that small values move right.
8: **until** done

Theorem 8.7. *Algorithm 8.6 sorts n values in $\sqrt{n}(\log n + 1)$ time in snake-like order.*

Proof. Since the algorithm is oblivious, we can use Lemma 8.4. We show that after a row and a column phase, half of the previously unsorted rows will be sorted. More formally, let us call a row with only 0's (or only 1's) *clean*, a row with 0's *and* 1's is *dirty*. At any stage, the rows of the mesh can be divided into three regions. In the north we have a region of all-0 rows, in the south all-1 rows, in the middle a region of dirty rows (possibly interspersed with clean rows). Initially all rows can be dirty. Since neither row nor column sort will touch the already clean rows in the northern and southern regions, we can concentrate on the middle region containing the dirty rows.

First we run an odd phase. Then, in the even phase, deviating from the algorithm description, let's run a peculiar column sorter: We group two consecutive rows in the middle region into pairs. Since odd and even rows are sorted in opposite directions, two consecutive rows look as follows:

$$00000 \ldots 11111$$

$$11111 \ldots 00000$$

Such a pair can be in one of three states. Either we have more 0's than 1's (in both rows together), or more 1's than 0's, or an equal number of 0's and 1's. Column-sorting each pair will give us at least one clean row (and two clean rows if "$|0| = |1|$"). Then we move the cleaned rows north/south and the middle region containing the dirty rows will be (roughly) halved in size.

What does this peculiar column sorter have to do with our algorithm? Well, a close look reveals that any column sorter sorts the columns in exactly the same way (we are very grateful to have Lemma 8.4!). Hence, we actually described what our algorithm does in the even phase.

All in all we need $2 \log m = \log n$ phases to remain with (at most) 1 dirty row in the middle which will be sorted (not cleaned) with the last row-sort. □

Remarks:

- There are algorithms that sort in $3m + o(m)$ time on an m by m mesh (by diving the mesh into smaller blocks). This is asymptotically optimal, since a value might need to move $2m$ times.
- Such a $\sqrt{n}$-sorter is cute, but we are more ambitious. There are non-distributed sorting algorithms such as quicksort, heapsort, or merge-sort that sort n values in (expected) $\mathcal{O}(n \log n)$ time. Using our n-fold parallelism effectively we might therefore hope for a distributed sorting algorithm that sorts in time $\mathcal{O}(\log n)$!

8.2 Sorting Networks

In this section we construct a graph topology which is carefully manufactured for sorting. This is a deviation from previous chapters where we always had to work with the topology that was given to us. In many application areas (e.g. peer-to-peer networks, communication switches, systolic hardware) it is indeed possible (in fact, crucial!) that an engineer can build the topology best suited for her application.

Definition 8.8 (Sorting Networks)**.** *A* comparator *is a device with two inputs x, y and two outputs x', y' such that $x' = \min(x, y)$ and $y' = \max(x, y)$. We construct so-called* comparison networks *that consist of wires that connect comparators (the output port of a comparator is sent to an input port of another comparator). Some wires are not connected to comparator outputs (we call them input wires), and some are not connected to comparator inputs (we call them output wires). A* sorting network *with* width n *has n input wires and n output wires. A sorting network routes n values given on the input wires through the wires and comparators of the network such that the values are sorted on the output wires.*

Remarks:

- Often we will draw all the wires on n horizontal lines, where n is the width of the network. Comparators are then vertically connecting two of these lines. An example sorting network is depicted in Figure 8.9.

Definition 8.10 (Depth)**.** *The depth of an input wire is* 0*. The depth of a comparator is the maximum depth of its input wires plus one. The depth of an output wire of a comparator is the depth of the comparator. The depth of a comparison network is the maximum depth (of an output wire).*

Remarks:

- The odd/even sorter explained in Algorithm 8.3 can also be described as a sorting network. An odd/even sorting network with width n has depth n.
- Note that a sorting network is an oblivious comparison-exchange network. Consequently we can apply Lemma 8.4 throughout this section.

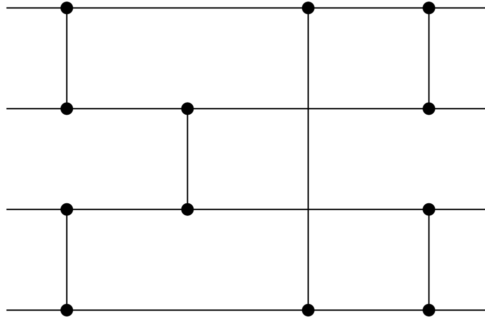

Figure 8.9: A sorting network with width 4, 6 comparators, and 16 wires.

Definition 8.11 (Bitonic Sequence)**.** *A* bitonic sequence *is a sequence of numbers that first monotonically increases, and then monotonically decreases, or vice versa.*

Remarks:

- $< 1, 4, 6, 8, 3, 2 >$ or $< 5, 3, 2, 1, 4, 8 >$ are bitonic sequences.
- $< 9, 6, 2, 3, 5, 4 >$ or $< 7, 4, 2, 5, 9, 8 >$ are not bitonic.
- Since we restrict ourselves to 0's and 1's (Lemma 8.4), bitonic sequences have the form $0^i 1^j 0^k$ or $1^i 0^j 1^k$ for $i, j, k \geq 0$.

Algorithm 8.12 Half Cleaner

1: A half cleaner is a comparison network of depth 1, where we compare wire i with wire $i + n/2$ for $i = 1, \ldots, n/2$ (we assume n to be even).

Lemma 8.13. *Feeding a bitonic sequence into a half cleaner (Algorithm 8.12), the half cleaner cleans (makes all-*0 *or all-*1*) either the upper or the lower half of the n wires. The other half is bitonic.*

Proof. Assume that the input is of the form $0^i 1^j 0^k$ for $i, j, k \geq 0$. If the midpoint falls into the 0's, the input is already clean/bitonic and will stay so. If the midpoint falls into the 1's the half cleaner acts as Shearsort with two adjacent rows, exactly as in the proof of Theorem 8.7. The case $1^i 0^j 1^k$ is symmetric. □

Lemma 8.15. *A bitonic sequence sorter (Algorithm 8.14) of width n sorts bitonic sequences. It has depth* $\log n$.

Proof. The proof follows directly from Algorithm 8.14 and Lemma 8.13. □

Algorithm 8.14 Bitonic Sequence Sorter

1: A bitonic sequence sorter of width n (where width is the number of input wires and we assume n to be a power of 2) consists of a half cleaner of width n, and then two bitonic sequence sorters of width $n/2$ each.

2: A bitonic sequence sorter of width 1 is empty.

Remarks:

- Clearly we want to sort arbitrary and not only bitonic sequences! To do this we need one more concept, merging networks.

Algorithm 8.16 Merging Network

1: A merging network of width n is a merger of width n followed by two bitonic sequence sorters of width $n/2$. A merger is a depth-one network where we compare wire i with wire $n - i + 1$, for $i = 1, \ldots, n/2$.

Remarks:

- Note that a merging network is a bitonic sequence sorter where we replace the (first) half-cleaner by a merger.

Lemma 8.17. *A merging network of width n (Algorithm 8.16) merges two sorted input sequences of length $n/2$ each into one sorted sequence of length n.*

Proof. We have two sorted input sequences. Essentially, a merger does to two sorted sequences what a half cleaner does to a bitonic sequence, since the lower part of the input is reversed. In other words, we can use the same argument as in Theorem 8.7 and Lemma 8.13: Again, after the merger step either the upper or the lower half is clean, the other is bitonic. The bitonic sequence sorters complete sorting. □

Remarks:

- How do you sort n values when you are able to merge two sorted sequences of size $n/2$? Piece of cake, just apply the merger recursively.

Algorithm 8.18 Batcher's "Bitonic" Sorting Network

1: A batcher sorting network of width n consists of two batcher sorting networks of width $n/2$ followed by a merging network of width n. (See Figure 8.19.)

2: A batcher sorting network of width 1 is empty.

Theorem 8.20. *A sorting network (Algorithm 8.18) sorts an arbitrary sequence of n values. It has depth* $\mathcal{O}(\log^2 n)$.

Proof. Correctness is immediate: at recursive stage k ($k = 1, 2, 3, \ldots, \log n$) we merge 2^k sorted sequences into 2^{k-1} sorted sequences. The depth $d(n)$ of the sorting network of width n is the depth of a sorting network of width $n/2$ plus

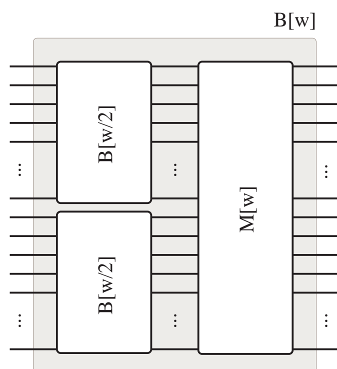

Figure 8.19: A batcher sorting network

the depth $m(n)$ of a merging network of width n. The depth of a sorter of level 1 is 0 since the sorter is empty. Since a merging network of width n has the same depth as a bitonic sequence sorter of width n, we know by Lemma 8.15 that $m(n) = \log n$. This gives a recursive formula for $d(n)$ which solves to $d(n) = \frac{1}{2}\log^2 n + \frac{1}{2}\log n$. □

Remarks:

- Simulating Batcher's sorting network on an ordinary sequential computer takes time $\mathcal{O}(n \log^2 n)$. As said, there are sequential sorting algorithms that sort in asymptotically optimal time $\mathcal{O}(n \log n)$. So a natural question is whether there is a sorting network with depth $\mathcal{O}(\log n)$. Such a network would have some remarkable advantages over sequential asymptotically optimal sorting algorithms such as heapsort. Apart from being highly parallel, it would be completely oblivious, and as such perfectly suited for a fast hardware solution. In 1983, Ajtai, Komlos, and Szemeredi presented a celebrated $\mathcal{O}(\log n)$ depth sorting network. (Unlike Batcher's sorting network the constant hidden in the big-O of the "AKS" sorting network is too large to be practical, however.)

- It can be shown that Batcher's sorting network and similarly others can be simulated by a Butterfly network and other hypercubic networks, see next chapter.

- What if a sorting network is asynchronous?!? Clearly, using a synchronizer we can still sort, but it is also possible to use it for something else. Check out the next section!

8.3 Counting Networks

In this section we address distributed counting, a distributed service which can for instance be used for load balancing.

Definition 8.21 (Distributed Counting)**.** *A* distributed counter *is a variable that is common to all processors in a system and that supports an atomic* test-and-increment *operation. The operation delivers the system's counter value to the requesting processor and increments it.*

Remarks:

- A naive distributed counter stores the system's counter value with a distinguished central node. When other nodes initiate the test-and-increment operation, they send a request message to the central node and in turn receive a reply message with the current counter value. However, with a large number of nodes operating on the distributed counter, the central processor will become a bottleneck. There will be a congestion of request messages at the central processor, in other words, the system will not scale.

- Is a scalable implementation (without any kind of bottleneck) of such a distributed counter possible, or is distributed counting a problem which is inherently centralized?!?

- Distributed counting could for instance be used to implement a load balancing infrastructure, i.e., by sending the job with counter value i (modulo n) to server i (out of n possible servers).

Definition 8.22 (Balancer)**.** *A balancer is an asynchronous device which forwards messages that arrive on the left side to the wires on the right, the first to the upper, the second to the lower, the third to the upper, and so on.*

Remarks:

- In electronics, a balancer is called a flip-flop.

Algorithm 8.23 Bitonic Counting Network.

1: Take Batcher's bitonic sorting network of width w and replace all the comparators with balancers.
2: When a node wants to count, it sends a message to an arbitrary input wire.
3: The message is then routed through the network, following the rules of the asynchronous balancers.
4: Each output wire is completed with a "mini-counter."
5: The mini-counter of wire k replies the value "$k + i \cdot w$" to the initiator of the ith message it receives (starting with $i = 0$).

Definition 8.24 (Step Property)**.** *A sequence $y_0, y_1, \ldots, y_{w-1}$ is said to have the* step property, *if $0 \le y_i - y_j \le 1$, for any $i < j$.*

Remarks:

- If the output wires have the step property, then with r requests, exactly the values $1, \ldots, r$ will be assigned by the mini-counters. All we need to show is that the counting network has the step property. For that we need some additional facts...

Facts 8.25. *For a balancer, we denote the number of consumed messages on the ith input wire by x_i, $i = 0, 1$. Similarly, we denote the number of sent messages on the ith output wire by y_i, $i = 0, 1$. A balancer has these properties:*

(1) A balancer does not generate output-messages; that is, $x_0 + x_1 \geq y_0 + y_1$ in any state.

(2) Every incoming message is eventually forwarded. In other words, if we are in a quiescent state (no message in transit), then $x_0 + x_1 = y_0 + y_1$.

(3) The number of messages sent to the upper output wire is at most one higher than the number of messages sent to the lower output wire: in any state $y_0 = \lceil (y_0 + y_1)/2 \rceil$ (thus $y_1 = \lfloor (y_0 + y_1)/2 \rfloor$).

Facts 8.26. *If a sequence $y_0, y_1, \ldots, y_{w-1}$ has the step property,*

(1) then all its subsequences have the step property.

(2) then its even and odd subsequences satisfy

$$\sum_{i=0}^{w/2-1} y_{2i} = \left\lceil \frac{1}{2} \sum_{i=0}^{w-1} y_i \right\rceil \text{ and } \sum_{i=0}^{w/2-1} y_{2i+1} = \left\lfloor \frac{1}{2} \sum_{i=0}^{w-1} y_i \right\rfloor.$$

Facts 8.27. *If two sequences $x_0, x_1, \ldots, x_{w-1}$ and $y_0, y_1, \ldots, y_{w-1}$ have the step property,*

(1) and $\sum_{i=0}^{w-1} x_i = \sum_{i=0}^{w-1} y_i$, then $x_i = y_i$ for $i = 0, \ldots, w-1$.

(2) and $\sum_{i=0}^{w-1} x_i = \sum_{i=0}^{w-1} y_i + 1$, then there exists a unique j ($j = 0, 1, \ldots, w-1$) such that $x_j = y_j + 1$, and $x_i = y_i$ for $i = 0, \ldots, w-1$, $i \neq j$.

Remarks:

- An alternative representation of Batcher's network has been introduced in [AHS94]. It is isomorphic to Batcher's network, and relies on a Merger Network $M[w]$ which is defined inductively: $M[w]$ consists of two $M[w/2]$ networks (an upper and a lower one) whose output is fed to $w/2$ balancers/comparators. The upper network merges the even subsequence $x_0, x_2, \ldots, x_{w-2}$, while the lower network merges the odd subsequence $x_1, x_3, \ldots, x_{w-1}$. Call the outputs of these two $M[w/2]$, z and z' respectively. The final stage of the network combines z and z' by sending each pair of wires z_i and z'_i into a balancer whose outputs yield y_{2i} and y_{2i+1}.
- It is enough to prove that a merger network $M[w]$ preserves the step property.

Lemma 8.28. *Let $M[w]$ be a merger network of width w. In a quiescent state (i.e., there is no message in transit), if the inputs $x_0, x_1, \ldots, x_{w/2-1}$ resp. $x_{w/2}, x_{w/2+1}, \ldots, x_{w-1}$ have the step property, then the output $y_0, y_1, \ldots, y_{w-1}$ has the step property.*

Proof. By induction on the width w.

For $w = 2$: $M[2]$ is a balancer and a balancer's output has the step property (Fact 8.25.3).

For $w > 2$: Let z resp. z' be the output of the upper respectively lower $M[w/2]$ subnetwork. Since $x_0, x_1, \ldots, x_{w/2-1}$ and $x_{w/2}, x_{w/2+1}, \ldots, x_{w-1}$ both have the step property by assumption, their even and odd subsequences also have the step property (Fact 8.26.1). By induction hypothesis, the outputs of both $M[w/2]$ subnetworks have the step property. Let $Z := \sum_{i=0}^{w/2-1} z_i$ and $Z' := \sum_{i=0}^{w/2-1} z'_i$. From Fact 8.26.2 we conclude that $Z = \lceil \frac{1}{2} \sum_{i=0}^{w/2-1} x_i \rceil + \lfloor \frac{1}{2} \sum_{i=w/2}^{w-1} x_i \rfloor$ and $Z' = \lfloor \frac{1}{2} \sum_{i=0}^{w/2-1} x_i \rfloor + \lceil \frac{1}{2} \sum_{i=w/2}^{w-1} x_i \rceil$. Since $\lceil a \rceil + \lfloor b \rfloor$ and $\lfloor a \rfloor + \lceil b \rceil$ differ by at most 1 we know that Z and Z' differ by at most 1.

If $Z = Z'$, Fact 8.27.1 implies that $z_i = z'_i$ for $i = 0, \ldots, w/2-1$. Therefore, the output of $M[w]$ is $y_i = z_{\lfloor i/2 \rfloor}$ for $i = 0, \ldots, w-1$. Since $z_0, \ldots, z_{w/2-1}$ has the step property, so does the output of $M[w]$ and the lemma follows.

If Z and Z' differ by 1, Fact 8.27.2 implies that $z_i = z'_i$ for $i = 0, \ldots, w/2-1$, except that there is a unique j such that z_j and z'_j differ by 1. Let $l := \min(z_j, z'_j)$. Then, the output y_i (with $i < 2j$) is $l + 1$. The output y_i (with $i > 2j + 1$) is l. The outputs y_{2j} and y_{2j+1} are balanced by the final balancer resulting in $y_{2j} = l + 1$ and $y_{2j+1} = l$. Therefore $M[w]$ preserves the step property. □

A bitonic counting network is constructed to fulfill Lemma 8.28, i.e., the final output comes from a Merger whose upper and lower inputs are recursively merged. Therefore, the following theorem follows immediately.

Theorem 8.29 (Correctness)**.** *In a quiescent state, the w output wires of a bitonic counting network of width w have the step property.*

Remarks:

- Is every sorting network also a counting network? No. But surprisingly, the other direction is true!

Theorem 8.30 (Counting vs. Sorting)**.** *If a network is a counting network then it is also a sorting network, but not vice versa.*

Proof. There are sorting networks that are not counting networks (e.g. odd/even sort, or insertion sort). For the other direction, let C be a counting network and $I(C)$ be the isomorphic network, where every balancer is replaced by a comparator. Let $I(C)$ have an arbitrary input of 0's and 1's; that is, some of the input wires have a 0, all others have a 1. There is a message at C's i^{th} input wire if and only if $I(C)$'s i input wire is 0. Since C is a counting network, all messages are routed to the upper output wires. $I(C)$ is isomorphic to C, therefore a comparator in $I(C)$ will receive a 0 on its upper (lower) wire if and only if the corresponding balancer receives a message on its upper (lower) wire. Using an inductive argument, the 0's and 1's will be routed through $I(C)$

such that all 0's exit the network on the upper wires whereas all 1's exit the network on the lower wires. Applying Lemma 8.4 shows that $I(C)$ is a sorting network. □

Remarks:

- We claimed that the counting network is correct. However, it is only correct in a quiescent state.

Definition 8.31 (Linearizable)**.** *A system is* linearizable *if the order of the values assigned reflects the real-time order in which they were requested. More formally, if there is a pair of operations o_1, o_2, where operation o_1 terminates before operation o_2 starts, and the logical order is "o_2 before o_1", then a distributed system is not linearizable.*

Lemma 8.32 (Linearizability)**.** *The bitonic counting network is not linearizable.*

Proof. Consider the bitonic counting network with width 4 in Figure 8.33: Assume that two *inc* operations were initiated and the corresponding messages entered the network on wire 0 and 2 (both in light gray color). After having passed the second resp. the first balancer, these traversing messages "fall asleep"; In other words, both messages take unusually long time before they are received by the next balancer. Since we are in an asynchronous setting, this may be the case.

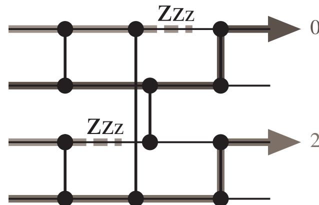

Figure 8.33: Linearizability Counter Example.

In the meantime, another *inc* operation (medium gray) is initiated and enters the network on the bottom wire. The message leaves the network on wire 2, and the *inc* operation is completed.

Strictly afterwards, another *inc* operation (dark gray) is initiated and enters the network on wire 1. After having passed all balancers, the message will leave the network wire 0. Finally (and not depicted in Figure 8.33), the two light gray messages reach the next balancer and will eventually leave the network on wires 1 resp. 3. Because the dark gray and the medium gray operation do conflict with Definition 8.31, the bitonic counting network is not linearizable. □

Remarks:

- Note that the example in Figure 8.33 behaves correctly in the quiescent state: Finally, exactly the values $0, 1, 2, 3$ are allotted.
- It has been shown that linearizability comes at a high price (the depth grows linearly with the width).

Chapter Notes

The technique used for the famous lower bound of comparison-based sequential sorting first appeared in [FJ59]. Comprehensive introductions to the vast field of sorting can certainly be found in [Knu73]. Knuth also presents the 0/1 principle in the context of sorting networks, supposedly as a special case of a theorem for decision trees of W. G. Bouricius, and includes a historic overview of sorting network research.

Using a rather complicated proof not based on the 0/1 principle, [Hab72] first presented and analyzed Odd/Even sort on arrays. Shearsort for grids first appeared in [SSS86] as a sorting algorithm both easy to implement and to prove correct. Later it was generalized to meshes with higher dimension in [SS89]. A bubble sort based algorithm is presented in [SI86]; it takes time $\mathcal{O}(\sqrt{n}\log n)$, but is fast in practice. Nevertheless, already [TK77] presented an asymptotically optimal algorithm for grid network which runs in $3n + O(n^{2/3}\log n)$ rounds for an $n \times n$ grid. A simpler algorithm was later found by [SS86] using $3n+O(n^{3/4})$ rounds.

Batcher presents his famous $\mathcal{O}(\log^2 n)$ depth sorting network in [Bat68]. It took until [AKS83] to find a sorting network with asymptotically optimal depth $\mathcal{O}(\log n)$. Unfortunately, the constants hidden in the big-O-notation render it rather impractical.

The notion of counting networks was introduced in [AHS91], and shortly afterward the notion of linearizability was studied by [HSW91]. Follow-up work in [AHS94] presents bitonic counting networks and studies contention in the counting network. An overview of research on counting networks can be found in [BH98].

Bibliography

[AHS91] James Aspnes, Maurice Herlihy, and Nir Shavit. Counting networks and multi-processor coordination. In *Proceedings of the twenty-third annual ACM symposium on Theory of computing*, STOC '91, pages 348–358, New York, NY, USA, 1991. ACM.

[AHS94] James Aspnes, Maurice Herlihy, and Nir Shavit. Counting networks. *J. ACM*, 41(5):1020–1048, September 1994.

[AKS83] Miklos Ajtai, Janos Komlós, and Endre Szemerédi. An 0(n log n) sorting network. In *Proceedings of the fifteenth annual ACM symposium on Theory of computing*, STOC '83, pages 1–9, New York, NY, USA, 1983. ACM.

[Bat68] Kenneth E. Batcher. Sorting networks and their applications. In *Proceedings of the April 30–May 2, 1968, spring joint computer conference*, AFIPS '68 (Spring), pages 307–314, New York, NY, USA, 1968. ACM.

[BH98] Costas Busch and Maurice Herlihy. A Survey on Counting Networks. In *WDAS*, pages 13–20, 1998.

[FJ59] Lester R. Ford and Selmer M. Johnson. A Tournament Problem. *The American Mathematical Monthly*, 66(5):pp. 387–389, 1959.

[Hab72] Nico Habermann. Parallel neighbor-sort (or the glory of the induction principle). Paper 2087, Carnegie Mellon University - Computer Science Departement, 1972.

[HSW91] M. Herlihy, N. Shavit, and O. Waarts. Low contention linearizable counting. In *Foundations of Computer Science, 1991. Proceedings., 32nd Annual Symposium on*, pages 526–535, oct 1991.

[Knu73] Donald E. Knuth. *The Art of Computer Programming, Volume III: Sorting and Searching*. Addison-Wesley, 1973.

[SI86] Kazuhiro Sado and Yoshihide Igarashi. Some parallel sorts on a mesh-connected processor array and their time efficiency. *Journal of Parallel and Distributed Computing*, 3(3):398–410, 1986.

[SS86] Claus Peter Schnorr and Adi Shamir. An optimal sorting algorithm for mesh connected computers. In *Proceedings of the eighteenth annual ACM symposium on Theory of computing*, STOC '86, pages 255–263, New York, NY, USA, 1986. ACM.

[SS89] Isaac D. Scherson and Sandeep Sen. Parallel sorting in two-dimensional VLSI models of computation. *Computers, IEEE Transactions on*, 38(2):238–249, feb 1989.

[SSS86] Isaac Scherson, Sandeep Sen, and Adi Shamir. Shear sort – A true two-dimensional sorting technique for VLSI networks. *1986 International Conference on Parallel Processing*, 1986.

[TK77] Clark David Thompson and Hsiang Tsung Kung. Sorting on a mesh-connected parallel computer. *Commun. ACM*, 20(4):263–271, April 1977.